DUNE COMPUTING

H. SCHELLMAN (OREGON STATE)
FOR THE COMPUTING CONSORTIUM

CDR progress

- Major design documents being finalized
  - Frameworks requirements – DONE -> HSF
  - Hardware data base – In production
  - SAM replacement
    - Metadata catalog – prototype
    - Rucio – in progress
    - Data dispatcher – design
    - Work plan in place
  - DAQ-Offline interface requirements – draft document
  - Use cases
    - ProtoDUNE
    - Analysis

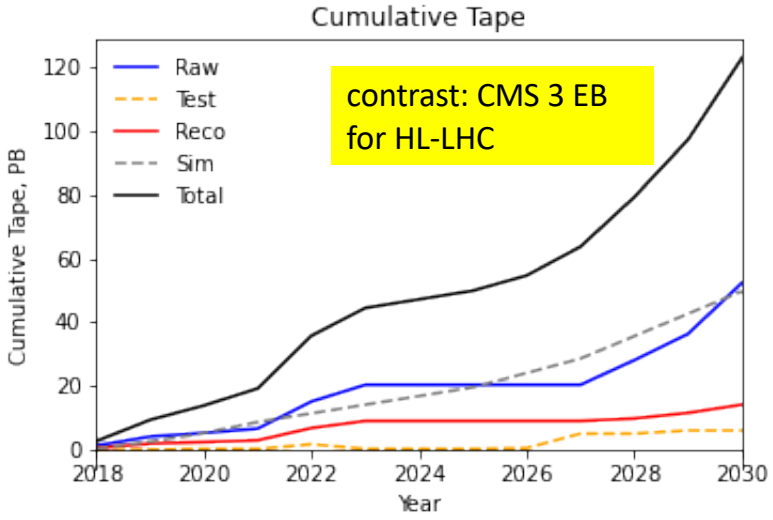
Two parts to Computing Infrastructure

Development and Operations

Institution	Country
York University	Canada
CERN	CERN
IN2P3	France
Edinburgh	UK
Manchester	UK
RAL/STFC	UK
Queen Mary Univ. London	UK
BNL	USA
Colorado State	USA
Fermilab	USA
LBNL	USA
MInnesota	USA
Oregon State University	USA
Wichita State	USA

Hardware contributions

Facility	Country
CBPF	BR
CA_Victoria	CA
CERN	CERN
FZU	CZ
PIC/CIEMAT	ES
CCIN2P3	FR
TIFR	IN
SURF/SARA	NL
JINR	RU
GridPP	UK
OSG	US
FNAL	US



contrast: CMS 3 EB for HL-LHC

CDR - Resource estimates to 2030

2 copies of raw data on tape

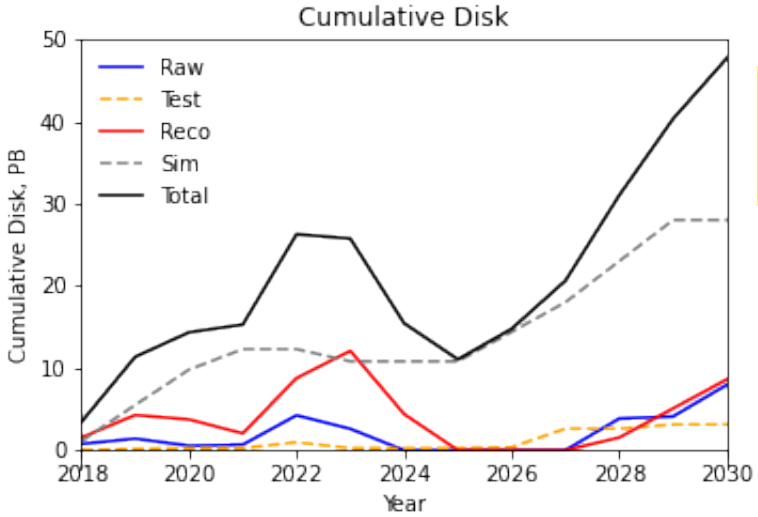
1 copy of "test" data stored for 6 months

1 copy of reco/sim on tape

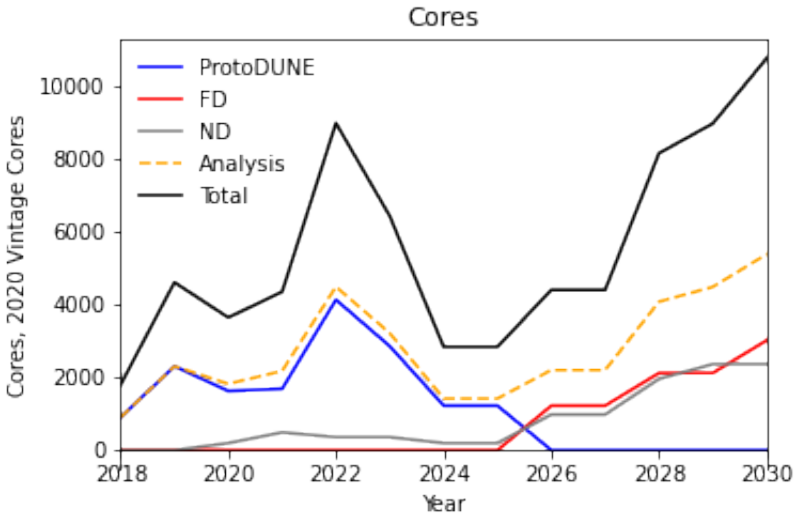
- Currently assume 2 reco passes and 1 sim pass/year
- Assume reco/sim resident on disk for 2 years

Assume 2 disk copies of reco and sim

- impose shorter lifetimes on tests and intermediate sim steps.

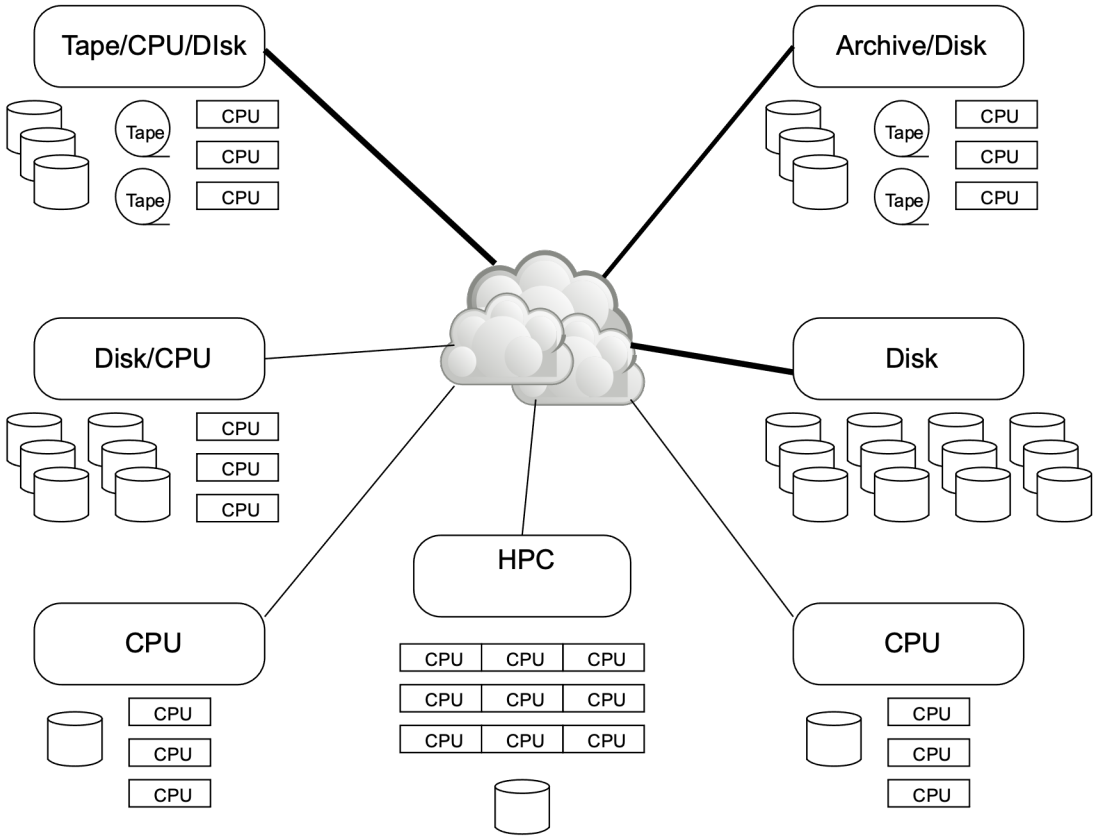


contrast: CMS 150 PB->2.2 EB in HL-LHC



CDR - Distributed computing model Does this work?

- Less “tiered” than current WLCG model
- Collaborating institutions (or groups of institutions) provide significant **services** (disk/CPU/archival)
- **Rucio** places multiple copies of datasets
- Workload/Data management system match data with appropriate delivery method
  - File already near local CPU
  - smart file location info
  - Direct copy to local cache
  - xrootd stream ← what we do now!
- Assumes good network connectivity
  - Currently working for 8,000 concurrent reconstruction jobs
  - Working with ESNET and European networking



Where we are now: Production pass 4

(re)processing ProtoDUNE-SP beam data

Process beam data and new simulation, starting on Nov. 9 –

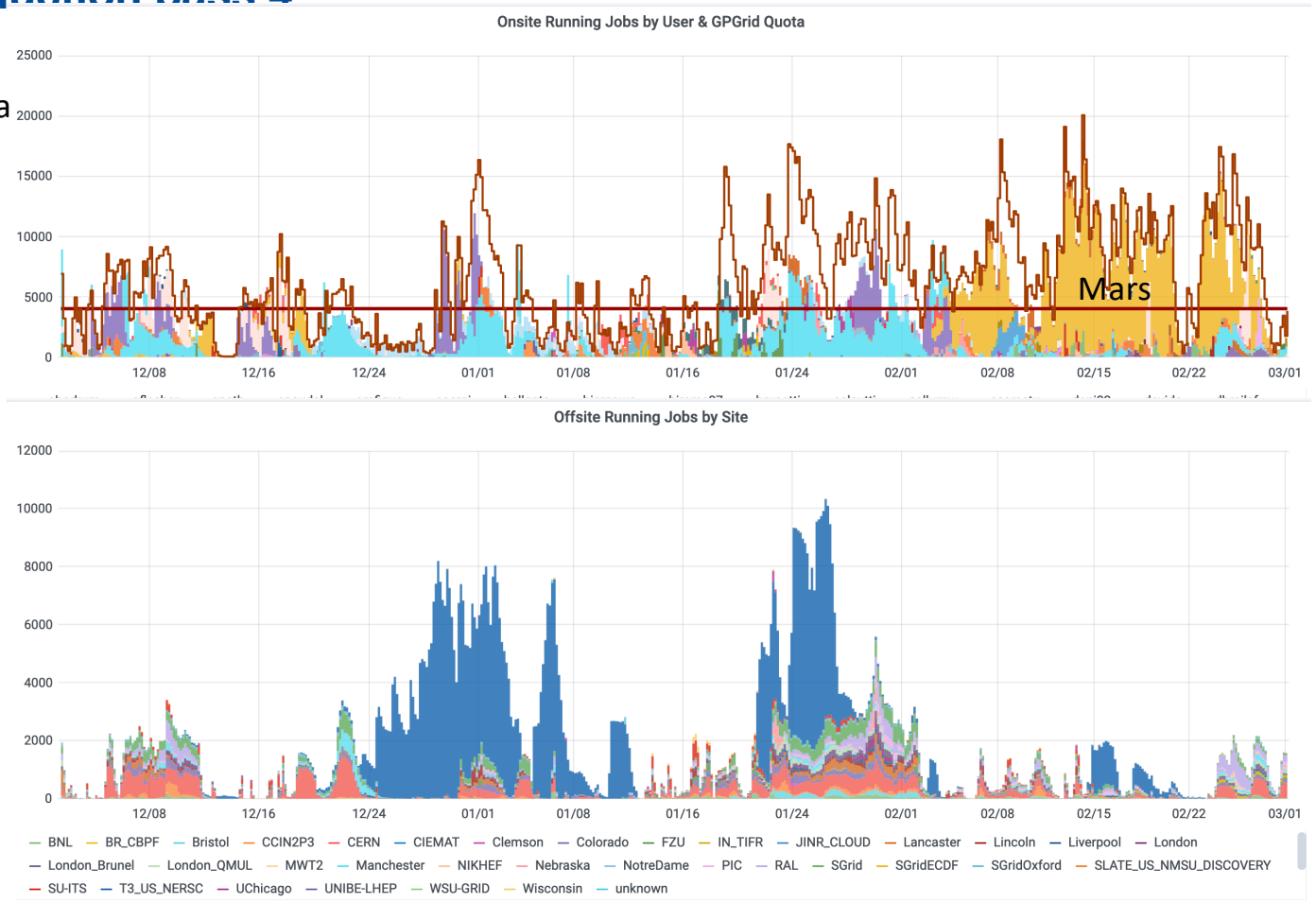
Simulation has multiple variations.

So far

Sim = 23M events (3 PB)
Data = 5.7 M events (300 TB)

Averaging 4000/4000 cores for on/offsite -- not that far off Full DUNE!

Memory for MC is >= 4.5GB

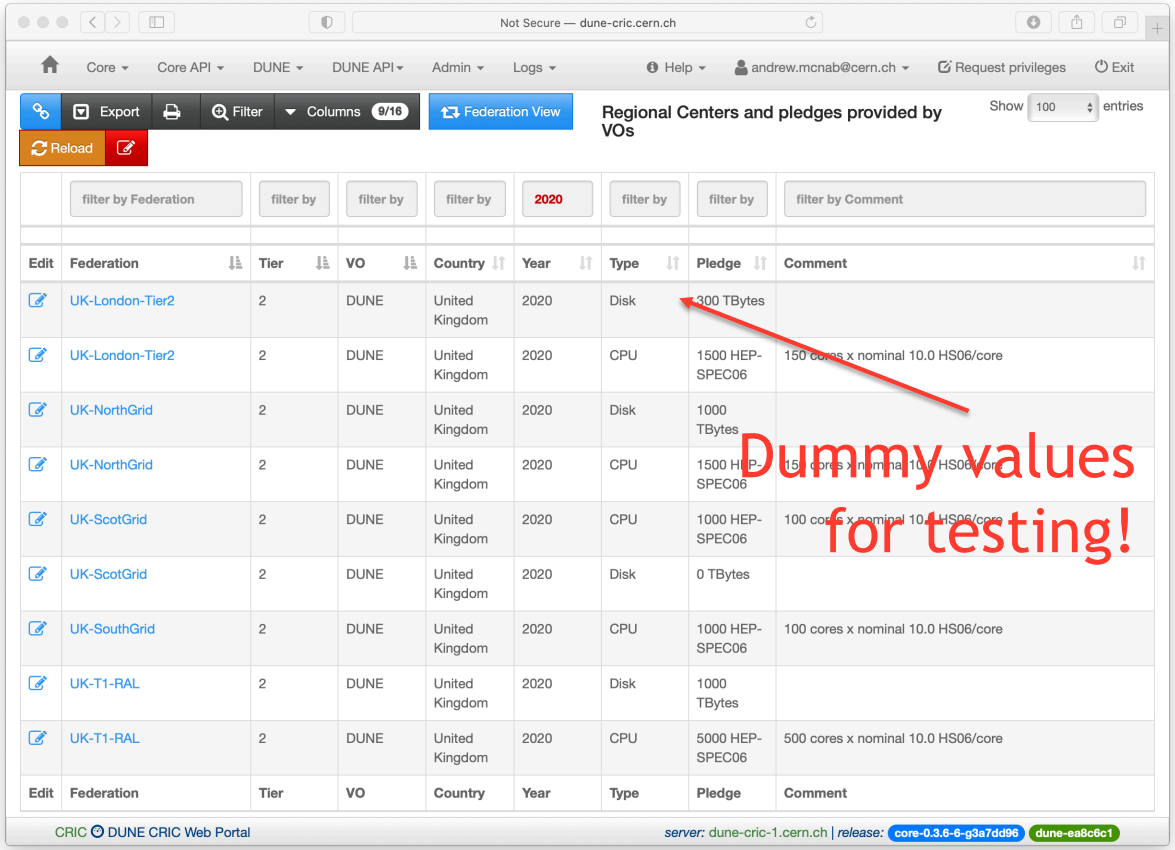


LBNF/DUNE

Responsibilities

- Tape storage
  - raw data – 2 copies – 1 at FNAL
  - sim/reco - 1 copy
- CPU
  - FNAL 25%, collaboration 75%
- Disk storage
  - National contributions 5-20% of the total from many countries
  - Pledges for 2021/2022 now being collected
- Network:
  - Working with ESNET on SURF->FNAL networking
  - Discussions with international partners (DUNEONE) on offshore compute network
  - Significant monitoring efforts underway

CRIC federation pledge submission

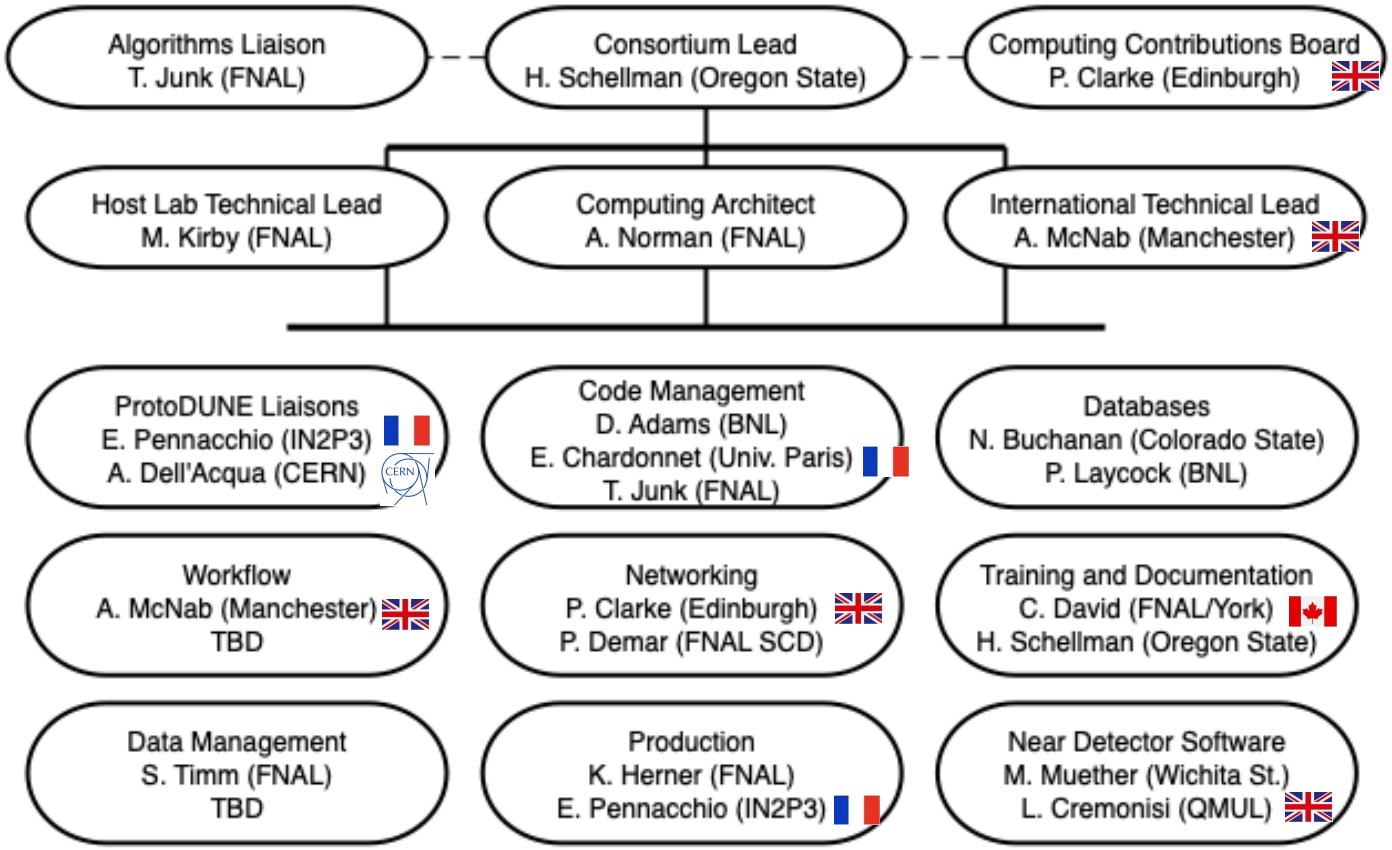


- DUNE using CRIC information system developed for WLCG
- Configuration details come from the OSG pilot factories
- Republishes this, and data gathered from other sources
- Used by DUNE ETF site monitoring to continually test site services
  - Tests new services as they come online, without any intervention by us
  - Intend to use CRIC as “The” canonical picture of the resources available to DUNE
- Now includes pledging

Conclusion on hardware resources

- We have identified and are formalizing resource contributions from collaborators worldwide.
  - CPU resources (looks good for protoDUNE)
    - shared resources from WLCG/OSG give us **lots of flexibility here**
  - disk resources (needs both contributions and code development)
    - **Disk needs to be DUNE specific**, will require substantial contributions from collaboration
  - tape resources (FNAL and CERN for now – will be come a big issue after 2030)
    - tight controls on data volumes and retention policies for intermediate steps.
  - networking – working with ESNET as part of their planning exercise
- **Need** to develop systems to monitor and use these resources
  - rest of talk

Development Organization



Development Scope

- Use common tools (ESNET, rucio, WLCG …) where possible
  - Downside: need to keep up as things change
  - Upside : we can make positive contributions to HEP infrastructure
- Detector is new
  - **New databases** need to be designed for conditions/calibrations
- DUNE events are very big and getting bigger 70 MB (PD) --> 3 GB (FD) → 185 TB (Supernova)
  - New framework
  - Memory management is … interesting …
  - HPC adaption
- Collaboration is large
  - Support (and train) large # of users
  - Need to monitor and coordinate large # of sites (32 already)
  - **support thousands of simultaneous connections to DB and data stores**
- Needs to be **ready** at small scale in **2022 for PD-II**, large scale between **2026-2029 for FD/ND**

Development: Frameworks task force has reported

Unique DUNE requirements:

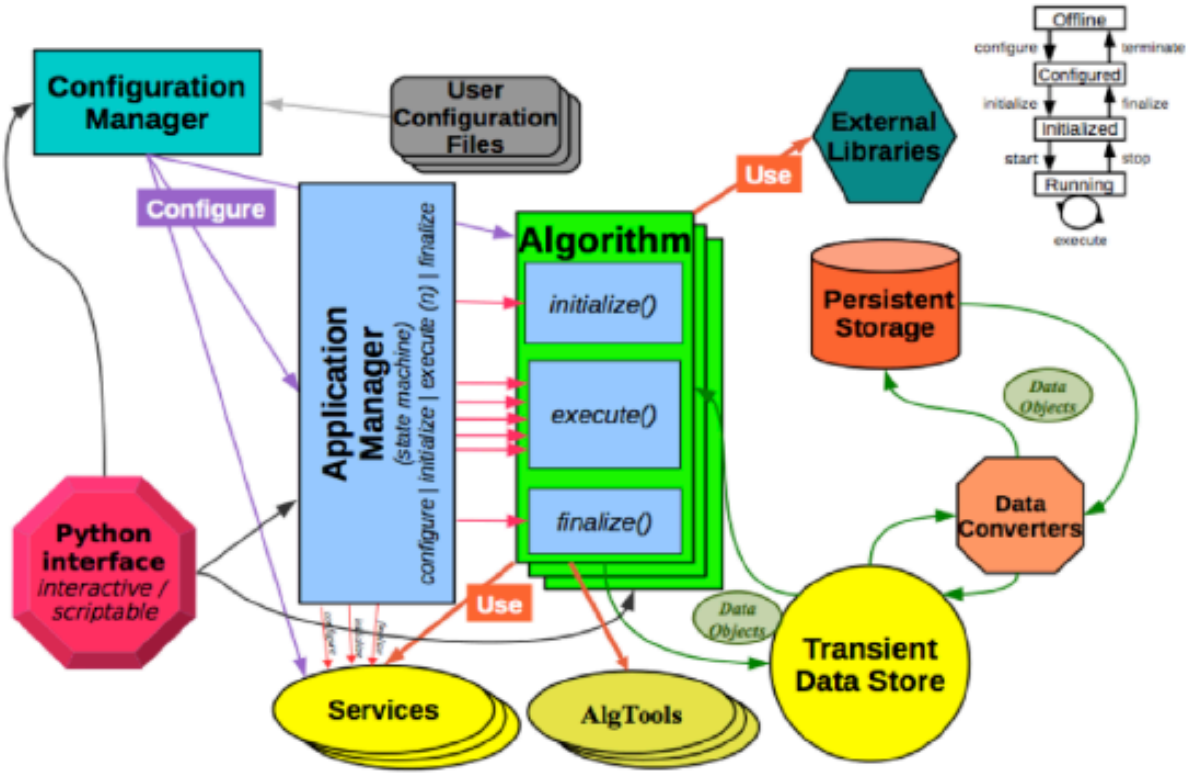
TPC/PD Data are simple on small scales → HPC

But 6 GB-100 TB readouts drive memory management requirements:

- Separate persistency/transient
- Precise tracking of provenance so parts can be reassembled
- multi-threading
- coherent processing across multiple architectures/sites

Can existing frameworks be modified to meet requirements?

Are reconstruction and analysis frameworks the same?



Final report

- Many useful comments and discussions, many thanks to the task force members and advisors
  - **DUNE members** - David Adams (BNL), Adam Aurisano (U. Cinc), Chris Backhouse (UCL), Mary Bishai (BNL), Claire David (York), Tom Junk (FNAL), Tom LeCompte (ANL), Chris Marshall (LBL), Brett Viren (BNL)
  - **Advisors** - Brian Bockelman (Madison), Chris Jones (FNAL), Kyle Knoepfel (FNAL), Liz Sexton-Kennedy (FNAL), Vakho Tsulaia (LBL), Peter Van Gemmeren (ANL)

- Converged on outstanding items and added an executive summary highlighting the **43 requirements** in broad categories:
  - Configuration requirements
  - Concurrency and Multithreading
  - Reproducibility and provenance
  - Random numbers, machine learning and conditions
  - Data and I/O layer
  - Memory management
  - Physics analysis

- *Report available here:*
- *https://docs.google.com/document/d/1OVR2d5kI_7auT3xbDuAMtQ9RG9SXo-AiH9SsyPopKuo/edit?usp=sharing*

Report is public
Executive summary -> CDR
Have asked HSF to evaluate the requirements

CMSSW/Art evolution?
Something else?

Development example: Databases

Now: Need substantial updates for PDUNE II to incorporate conditions/calibrations cleanly

- ✓ Beam database
- ✓ Hardware database for FD
  - Already needed yesterday
  - Going into production now….

2021-2023: ProtoDUNE-II run and analysis

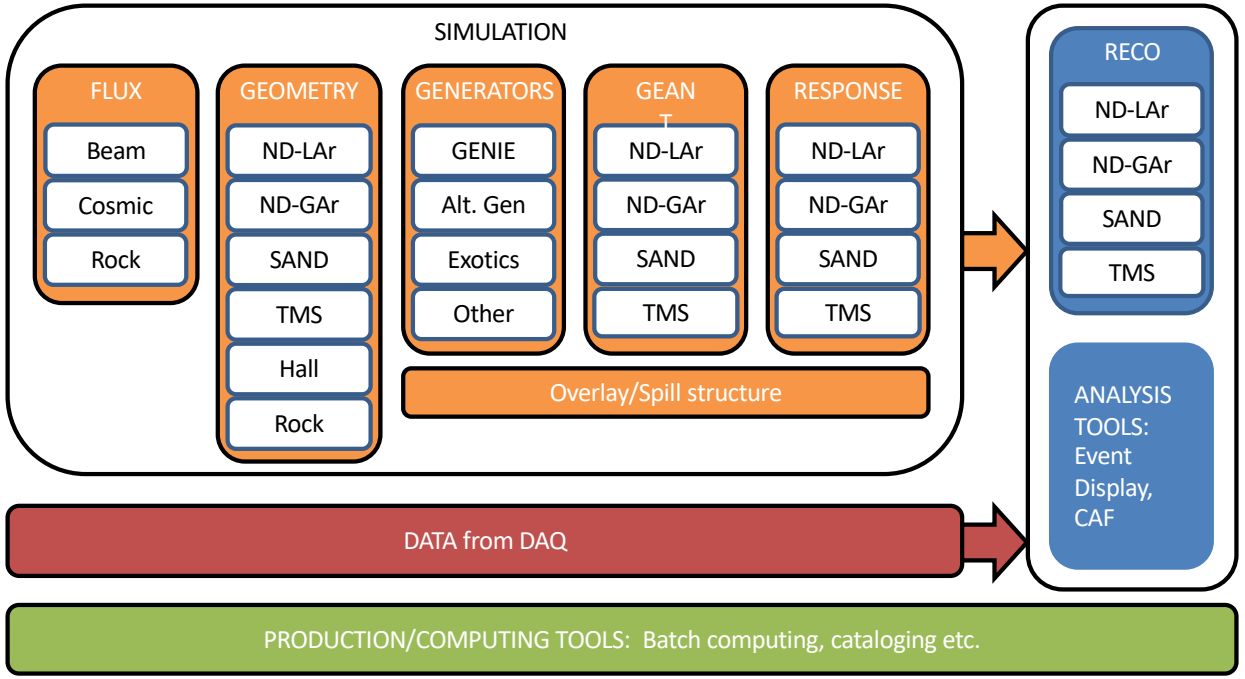
- ❑ Data Catalog
  - Developing more efficient MetaCat and Data tracking db's
- ❑ Compute systems monitoring
- ❑ Conditions/Slow controls/Calibrations

2024-2027: 2nd iteration to go to full scale for DUNE

- ❑ Calibration/Conditions at full scale for ND/FD
  - FD/ND will have many more channels than existing IF DB's were built to support
  - ~400,000 channels/FD module, many more for ND.
  - needs significant effort early on for design
  - will need significant horsepower to serve information to 10,000 cores worldwide.

Near Detector: Computing implications

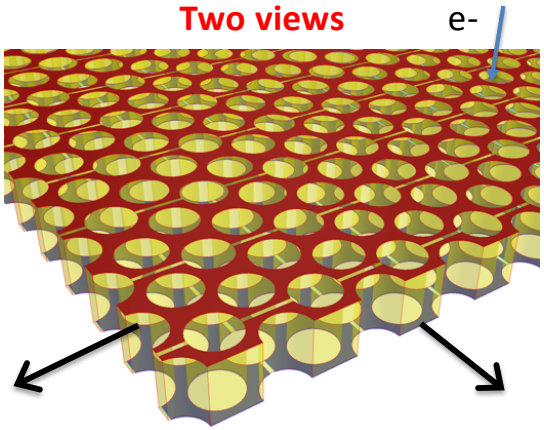
- Much more diverse than FD
- But much smaller events – can use more conventional computing methods
- Currently using standalone+ GArSoft
  - edep-sim
    - sand-fluka/sand-stt
    - Cube-recon
    - well documented but needs to move to full integrated framework
- Included in frameworks discussions – should drive some requirements
- Linda Cremonisi and Mathew Muether (and Tom Junk)



Vertical Drift technology implications for computing

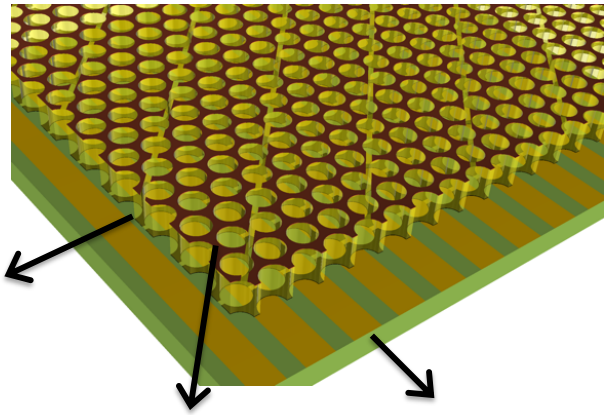
- Similar (or 32% larger for 3D) TPC channel count, smaller PD channel count
- Otherwise algorithmically similar to APA technology
- CRP readout already integrated into reconstruction chain.

Two views



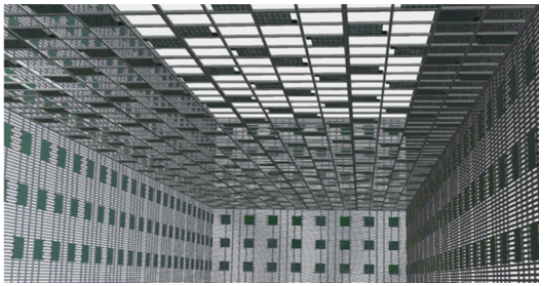
- ✓ Collection strips in the transverse direction, 5.2mm width
- ✓ Induction strips in the longitudinal direction, 5.2mm width
- ✓ 389'120 electronics channels

Three views



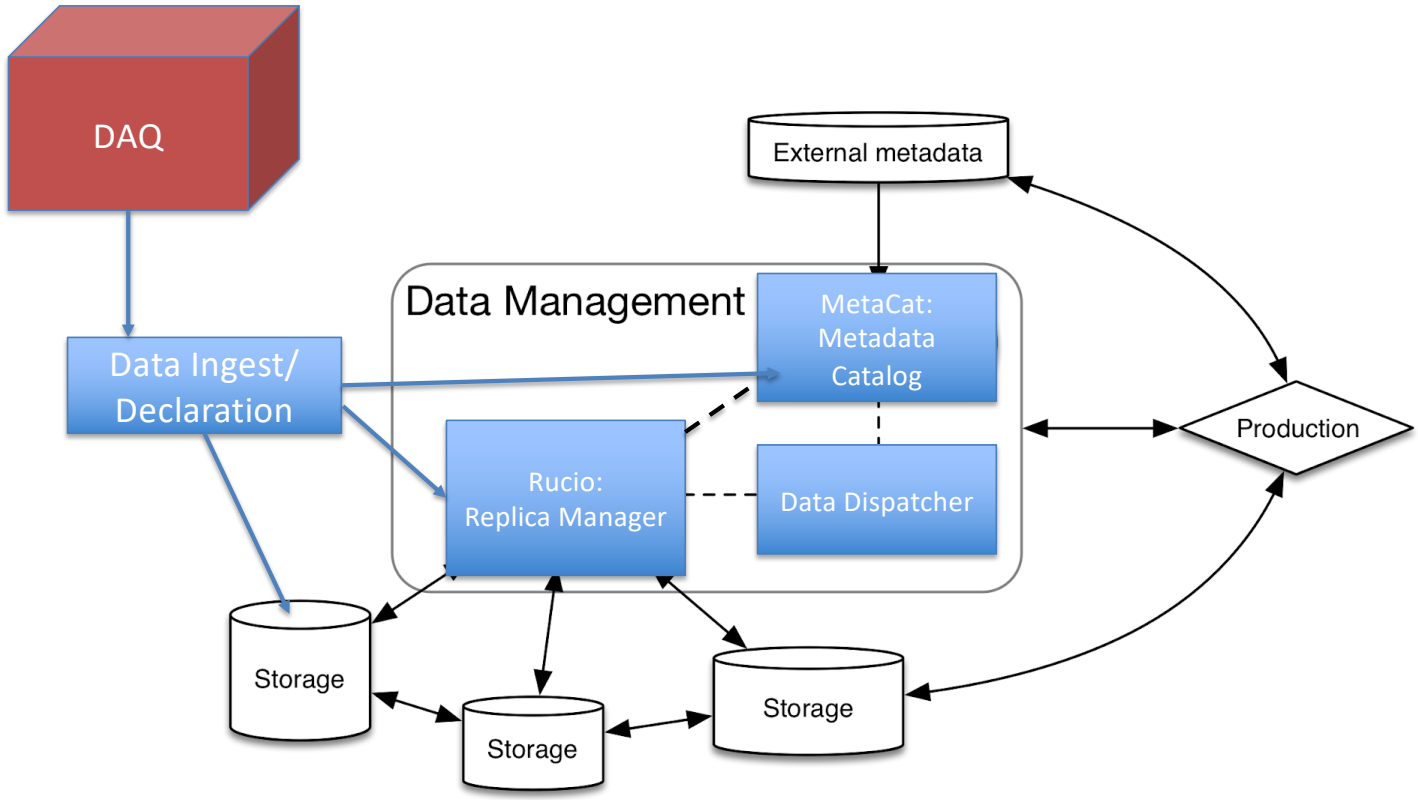
- ✓ Adding a 3th view at ~45°
- ✓ With strip width 8.7 mm
- ✓ ~32% more channels

Photon detectors



- ✓ Additional Arapuca units placed on the field cage, a 4π view
- ✓ It will improve the physics performance, even with respect to the HD detector

DUNE Data Management System

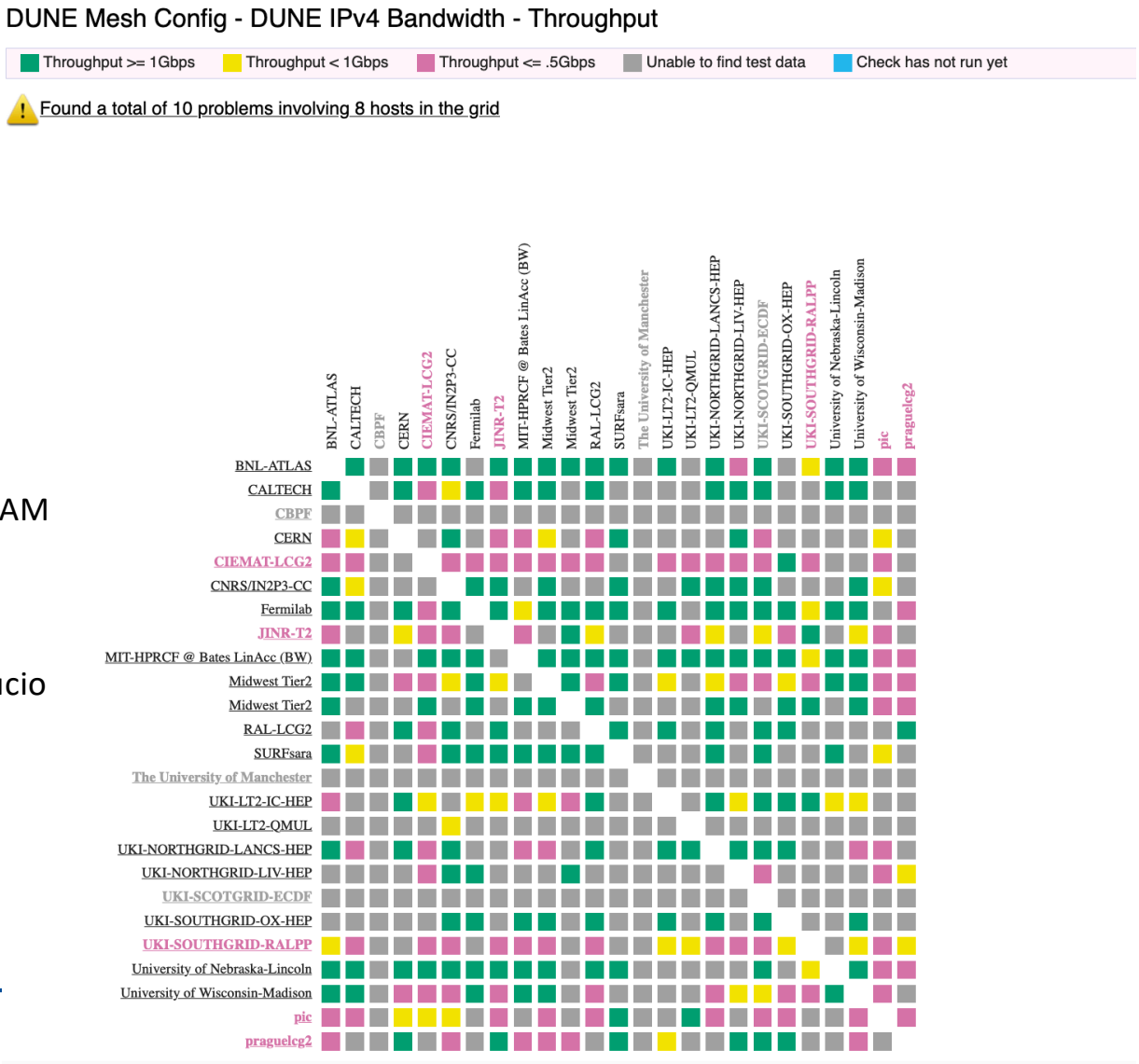


DUNE Data Management Projects

- **Development:**
  - **Rucio** Logical to Physical File mapping for Tape sites (non-deterministic) [James Perry, Edinburgh]
  - **MetaCat** moving towards deployment. [Igor Mandrichenko, FNAL]
  - **Data Dispatcher**—rework of SAM project functionality—[Brandon White, FNAL] Start summer 2021
  - **Data Ingest Daemon**—(Rework of FTS-Light functionality) - Start Fall 2021
  - **Data Transfer Daemon**--(Rework of FTS functionality to declare to Rucio/Metacat) – Start Fall 2021
- **Operations:**
  - Backloading all data into Rucio—[S. Timm, FNAL] – In progress
  - CTA testing @ CERN—[Wenlong Yuan, Edinburg] -- In progress
  - Rucio daily testing—[S. Timm, FNAL + Oregon State CS students] – In progress
  - Rucio transfer speed monitoring / Sam-Xrootd monitoring [Oregon State students]
  - Deployment of production OKD-based Rucio Server. [B. White, FNAL]

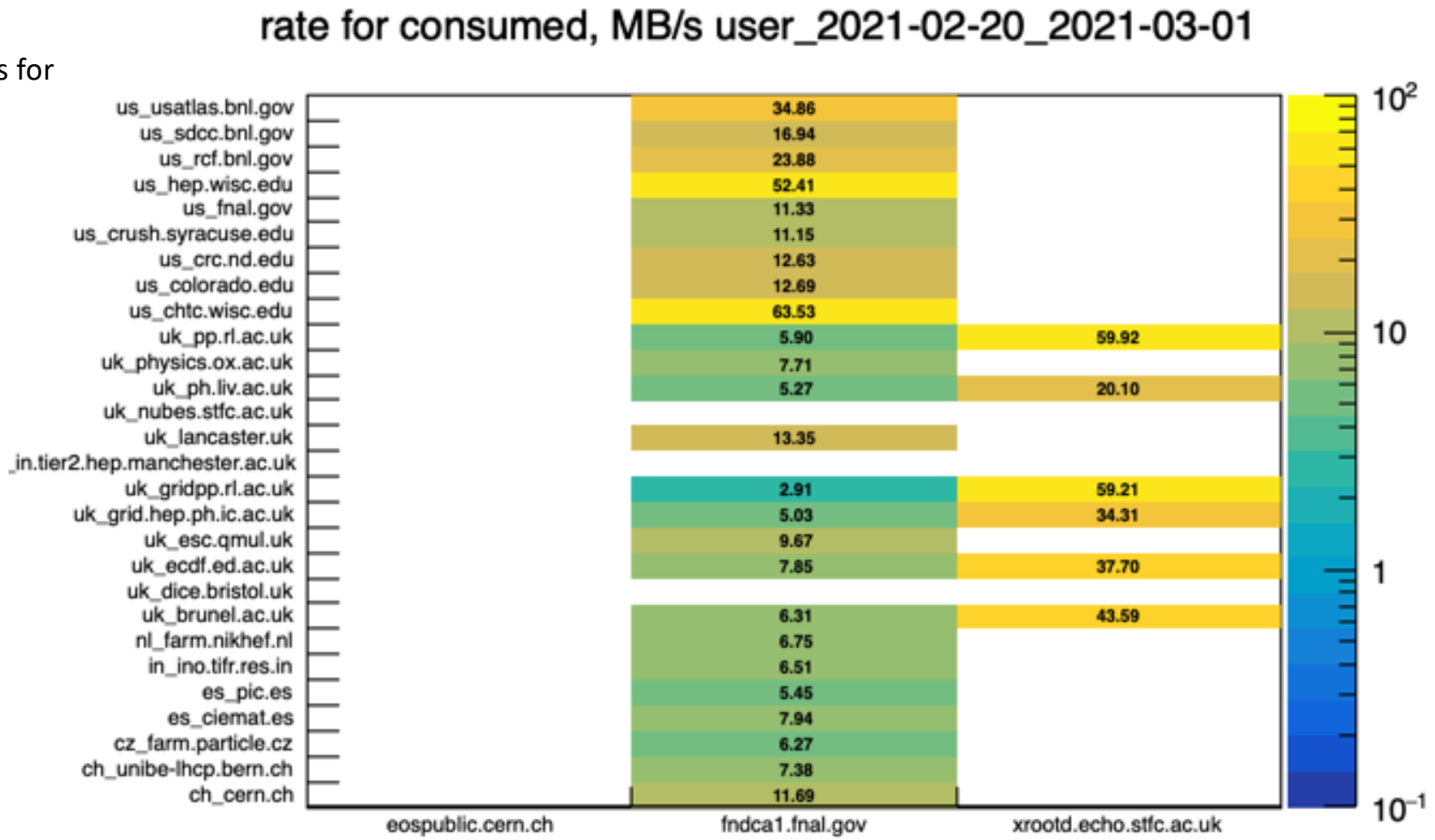
Monitoring

- perfsonar, Terry Froy, QMUL – working
- fifemon/POMS monitoring batch systems
- Development projects to mine rucio and SAM records for successful rucio transfers and xrootd streams
- Preliminary data show data moved with rucio to UK are now being used by users



xroot data rates for successful user processes

CPU

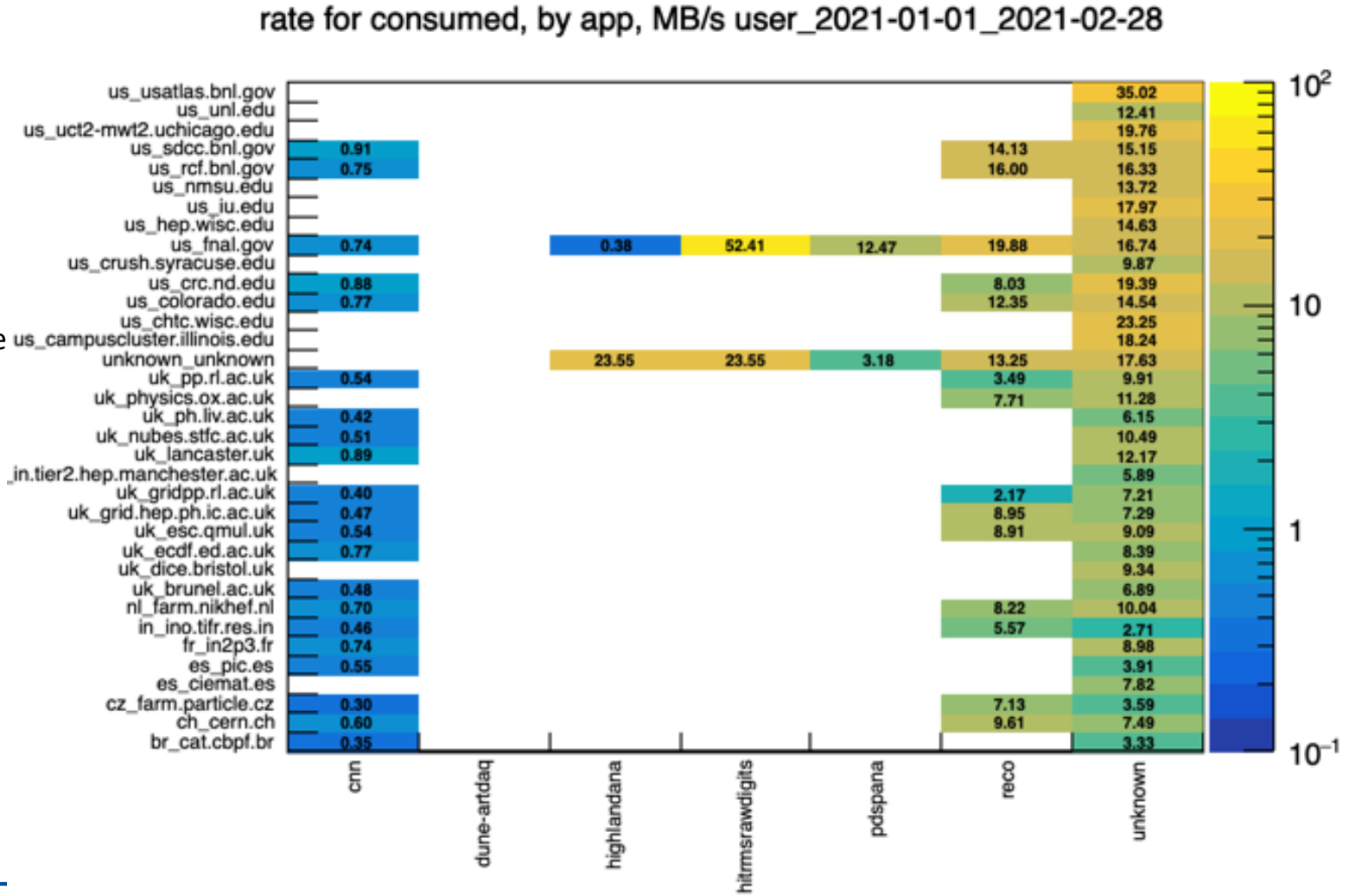


CERN only has raw data

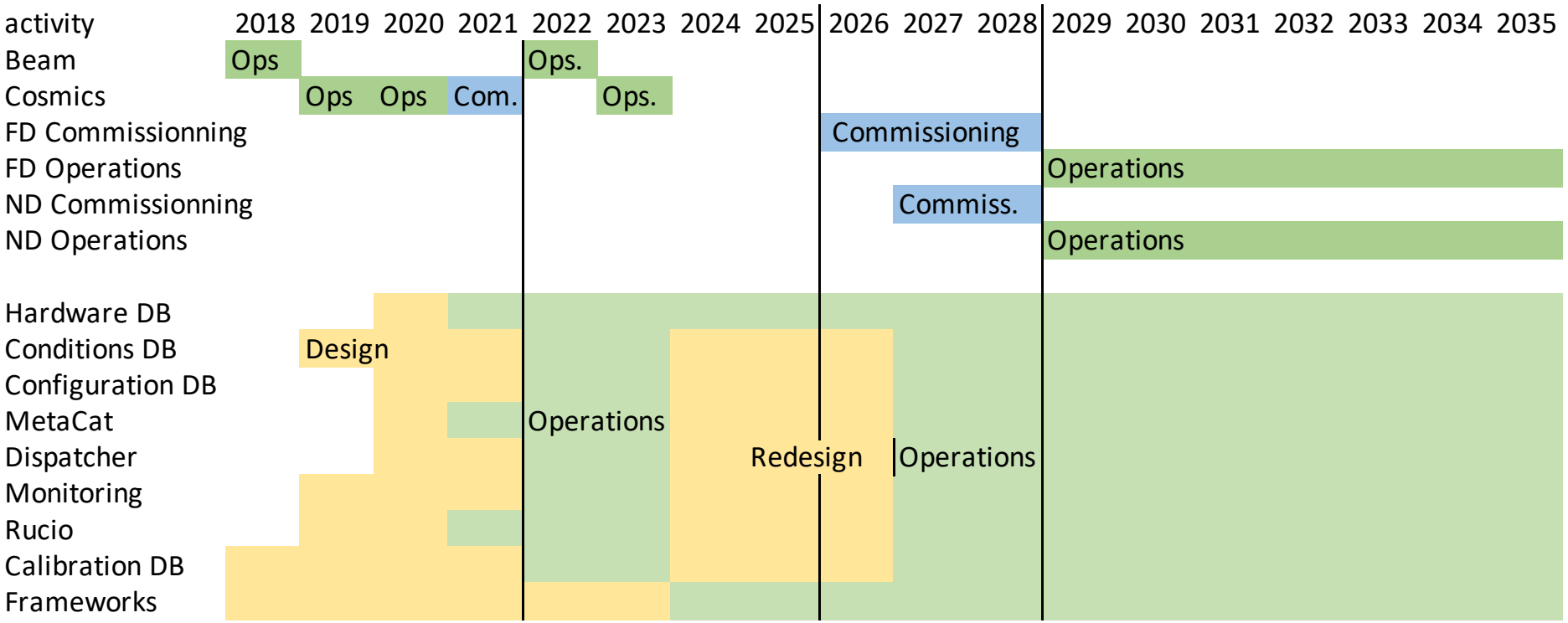
DISK location

What were users doing?

Use this to info to predict network use patterns.

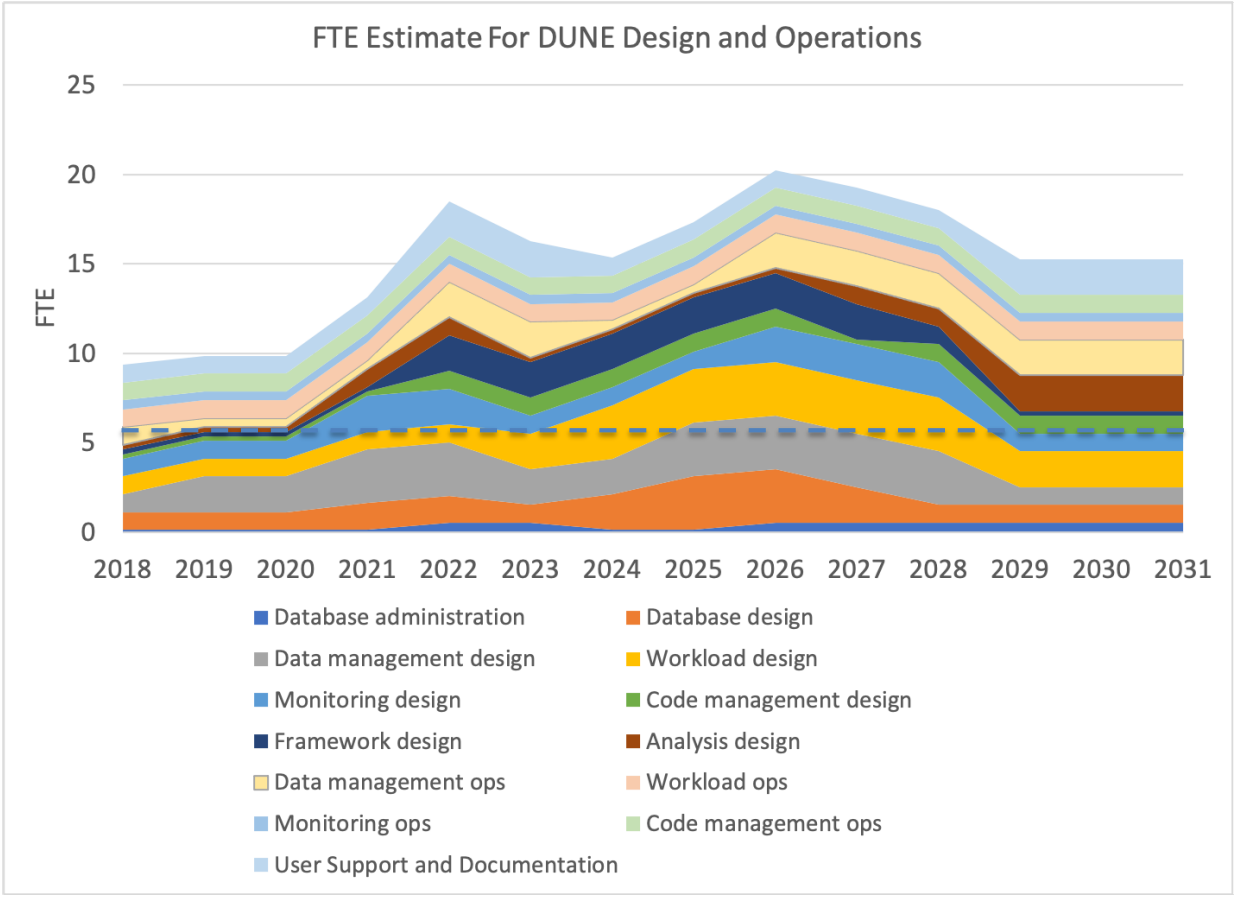


Development: Rough timeline



FTE estimate. Does not include shared facility (storage etc.) costs

- Some effort (mainly operations – pastels at top) can be trained collaboration physicists.
- Rest requires experts
- Currently have around 5 FTE experts (FNAL + collab), all in-kind contributions except UK DUNE funded personnel.
- Expert need is greatest for ProtoDUNE 2 and pre-operations in 2024-2028. 5-10 FTE > 50% US



Conclusions

Significant collaboration contributions to hardware and development effort have been identified

Storage contributions are high priority

More expert effort for development is needed for protoDUNE II in 2021-2022 and pre-operations starting in 2024-2027

Title	Link & platform
DUNE Software Framework Requirements Taskforce Report	GoogleDocs
Near Detector Data Model	Overleaf
Data Tracking	GoogleDocs
Metadata Catalog requirements	GoogleDocs
ESNET report	docdb-20816
Database description and definitions	Overleaf
Database hardware database requirements	Overleaf
Sites and Centres Model	GoogleDocs
Computing CDR	Overleaf
DAQ/Computing Metadata	GoogleDocs
MetaCat Documentation	html

Development contribution examples

- Not a complete list
- FNAL contributions are largely parts of shared projects across IF and EF:
  - LArSoft
  - Frameworks
  - Rucio integration
  - Storage systems
  - Generators
  - Networking
  - DAQ
  - Grid integration
  - Code management
  - < 5 FTE spread across many people paid through mainly through shared projects, physicist effort
- BNL
  - Data quality monitoring suite
  - Signal processing code
  - Code management
  - DB and framework leadership
- UK/GridPP
  - Rucio
  - data and workload management
  - monitoring systems
- IN2P3
  - Code management
  - Dual Phase software
  - Slow controls interface
- CERN
  - File transfer systems
  - ProtoDUNE operations
  - WLCG/common tools